END
4 84

AD A138787

A UNIFIED MODEL OF ATTENTION AND PROBLEM SOLVING

Earl Hunt and Marcy Lansman

February 1984

MAR 5 1984

A

84 03 02 007

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 84-1	2. GOVT ACCESSION NO. AD-A135 71?	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A UNIFIED MODEL OF ATTENTION AND PROBLEM SOLVING		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Earl Hunt and Marcy Lansman		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0631
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Washington Seattle, WA 98195		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (code 442PT) Arlington, VA 22217		12. REPORT DATE February, 1984
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Attention, problem solving, simulation, computer models, production systems, choice reaction time, stimulus repetition, Stroop, dual tasks, controlled processing, automatic processing.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The concept of production-executing machines has been used to construct a number of simulations of human problem solving. With a few exceptions the simulations have been of problem solving in situations in which there is no real-time pressure to respond. Such situations are typical of the paradigms used to study attention and performance. A model of production selection and execution has been developed that subsumes the previous problem solving models and that can be applied to real-time situations. The model has been

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

used to simulate human performance in choice reaction time, stimulus repetition, dual task, and cue-conflict (Stroop) situations. This extends the use of production system models to encompass both problem solving and attention limited behavior.

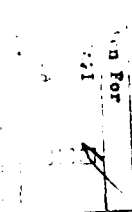A Unified Model of Attention and Problem Solving *

Earl Hunt

The University of Washington

and

Marcy Lansman

The University of North Carolina at Chapel Hill

Running head: ATTENTION AND PROBLEM SOLVING

The development of theories of attention and the study of problem solving represent distinct traditions in the study of cognition. In this paper we present a model of human performance that is intended to integrate theoretical concepts derived from the two fields. Two concepts are central to the model. One idea, taken from the study of problem solving, is the idea that thinking can be described as the activation of productions - pattern-action rules that constitute the elementary steps in problem solving. The second idea, taken from the study of memory and attention, is that memory records can be thought of as a set of interconnected nodes, called a semantic network. In a semantic network model, memories are aroused by the automatic spread of activity through the network. In our model a production is associated with each node in the semantic network, and that production is aroused when the corresponding node reaches a critical level of activity. The details of pattern recognition and production arousal processes are spelled out below. Because the model deals with the interaction between spreading activation and production execution, it will be referred to as the Production Activation Model.

The model has been implemented as a computer program.

This has made possible the simulation of a variety of phenomena from the literature on attention. Attentional phenomena were chosen for simulation because they depend upon the interaction between semantic activation and production execution. This contrasts with phenomena tht depend upon the characteristics of either the production system, alone, or the semantic network, alone. To illustrate, Newell and Simon's (1972) simulations of human problem solving were determined entirely by the characteristics of production systems they used. Their program did not contain an analog of a semantic network. On the other hand, Anderson (1983, a, b) reports a simulation of the "fan effect," the fact that the time taken to recall a fact about an object increases with the number of facts that are known. The simulation of the fan effect depended only upon the process of activation in a semantic network.

The decision to focus on the interaction between semantic activation and production execution does not mean that the resulting model is unable to handle phenomena that depend on either process alone. Since the model is capable of executing productions, it inherits the capability for simulating problem solving that has already been demonstrated for production system models in numerous

studies. (See, in particular Newell and Simon, 1972). Since it contains a semantic network, the model could be used to simulate phenomena produced by semantic activation. In computer terminology, the model can be looked upon as an interpreter of production systems. Therefore it can execute any simulation of problem solving that is stated as a production system program.

The attempt to deal with attentional and problem solving phenomena within the same model is challenging, because studies that focus on only one of these fields can afford to ignore problems that must be dealt with by the other. Consider problem solving. Previously reported production system models have been used to simulate the accuracy of problem solving in situations in which the subject was asked to perform a single task, without any constraints on attention. These studies dealt with neither the time pressures involved in problem solving, nor the effects of distracting concurrent stimulation. The production activation model deals with both. On the other hand, theories of the activation of memories rarely give a satisfactory description of the process by which increased activation is translated into action. A model in which the units of memory are productions rather than (ill-defined) engrams makes the connection between

activation and action explicit.

This paper is organized into three sections. The
first contains a description of the model and the
simulation program. The second presents the results of
simulations of some of the phenomena seen in popular
paradigms of the attention and performance literature.
The paper does not go into great detail in discussing any
one simulation, since the intent is to show the breadth of
the model rather than to explore its potential for depth
of explanation. (Subsequent, more specialized papers are
planned to discuss each of the simulations in more
detail.) The third section discusses the psychological
significance of the model, and compares it to other models
of mental action.

Description of the model

We assume that when a stimulus is presented two
concurrent information processing sequences are initiated.
One involves a pattern recognition process that culminates
when a label identifying the stimulus is placed in working
memory. The label provides an interpretation of the
stimulus that can serve as a trigger for further actions.
This sequence of pattern recognitions, actions, and

further pattern recognitions will be called the
"controlled" information processing sequence. In
addition, stimulus presentation is assumed to initate an
"automatic" processing sequence that behaves in quite a
different way. Instead of relying on pattern recognition
guided by information in working memory, the automatic
processing sequence relies upon the spread of activation
levels from one engram in memory to associates of that
engram, without involving the working memory system. In
the following section, these two systems are described in
detail.

The Controlled Processing System

As mentioned above, controlled processing can be
envisaged as the operation of a production system
interpreter (Newell, 1973). Such interpreters contain two
parts: a set of productions (pattern-action pairs) stored
in long term memory, and a blackboard area that contains
information about the current situation. At each cycle
the pattern parts of all productions are compared to
information on the blackboard. If any patterns are
recognized, one or more of the associated actions are
taken. Figure 1 shows the relationships between the
blackboard and long-term memory utilized in the Production

Activation Model. The blackboard itself is divided into three areas: two sets of external channels (one visual and one auditory) that contain information presented to the system, and a working memory area that contains information that the system itself generates as it interprets problem solving situations.

Figure 1 here

The information flow implicit in Figure 1 will now be explained. Information is placed in the external channels by the "environment," i.e., by a process outside the scope of the model itself.   The information presented may be in either an "auditory" or a "visual" code. These names have been chosen for their obvious analogy to sensory modalities, but computationally the only distinction between them is that the stimuli are described in different codes. Information in the external channels is examined by the productions available in long-term memory. When an auditory or visual pattern is recognized, an internal label for the stimulus is placed in the working memory area. The internal label may either be in the (auditory or visual) sensory code in which the stimulus was presented, or it may be in an internal "semantic"

code. If the label is in a sensory code, it is placed in either the "auditory" or "visual" channel of the working memory area. These channels provide a way for the system to respond to internally generated stimuli, represented as sensory codes. The internal channels are thus analagous to Baddeley's (1976) concept of auditory and visual buffers in working memory.

If the internal label is in the semantic code, it may be placed in any of the the several semantic channels in the working memory area. Baddeley and others have stressed the need for a modality free representation of information in working memory. The semantic code provides such a representation.

The productions in long-term memory continually are matched against both the external and working memory channels. Thus configurations of working memory may themselves serve as stimuli for further actions. For example, suppose that two stimuli were placed on separate external visual channels. The model could be "programmed" (i.e., provided with an appropriate production system) that would select one of them, place it in the visual channel of working memory, and then use the internal visual code as a stimulus to place a semantic

interpretations of the origin stimulus in the semantic *working memory area.*

The following example, which is based on an actual simulation study, illustrates the process in more detail. Consider a two-choice reaction time study in which either of two visual stimuli - Stimulus 1 or Stimulus 2 - can be presented. The subject's task is to identify the stimulus, by making Response 1 if Stimulus 1 has been presented and Response 2 if Stimulus 2 has been presented. A production system can be constructed *using two pairs of rules:*

1. If stimulus x (x =1,2) appears on a visual *channel, place the signal "recognized stimulus x" in* working memory. Two productions are required, one for each value of x.

2. If the signal "recognized stimulus x" is in working memory make response x.

The condition part of a rule (the "if" clause) will be referred to as the pattern of a production. The consequent part (the "then" clause) will be referred to as the action. In controlled processing, a sequence of

productions is executed in an order that is determined by their level of activation. The first step in controlled processing is determination of the extent to which there is *a match between the stimulus and each of the patterns* in long-term memory. A complete model of how stimuli and patterns are matched would constitute a theory of perception, which is quite beyond the scope of our work. Instead of including such a theory, the model includes a pattern recognition process that is proposed as sufficiently descriptive of human perception for our *purposes.*

The stimulus is represented as an ordered list of features, drawn from a feature alphabet, or code, appropriate to that particular type of stimulus. Distinct similarity matrices are associated with auditory, visual, and semantic codes. The (ij)th entry of this matrix is a number, between zero and one, indicating the extent to which the ith value of a feature in that code *resembles* the jth value. For example, if the stimuli were figures of varying shape, the entry for (triangle, circle) would be near zero, and the entry for (circle, ellipse) would be near one. The diagonal (ii) entries of the feature matrix are always one, indicating that a stimulus feature most resembles itself.

The similarity matrix *notation* provides a flexible way of describing features, since it allows for the possibility of a confusion without requiring that all conceivable confusions be permitted. For instance, a stimulus similarity matrix could be constructed to permit confusion between colors (e.g., red and orange) or forms (triangle and square) but never between colors and forms. *Psychologically* it would be more realistic to think of sub-dictionaries (and confusion matrices) within a large dictionary of *visual* or *acoustic features.* In practice, it is easier to maintain a single dictionary for each type of sensory code.

Patterns are defined by ordered lists of pairs, (f,w), where f is a feature in the appropriate code and w is an indicator of the importance of the feature to the pattern. The value of w may vary *from -1 to +1, depending on whether the feature* is contraindicated, irrelevant (w=0) or mandatory. The resemblance of a stimulus to a pattern is computed using Luce's (1956) choice rule,

$$(1) \quad \text{Resemblance of a stimulus to pattern} \quad \frac{\sum_{j=1}^{k} w(j) \, sim(s(j),p(j))}{\sum_{j=1}^{k} w(j)}$$

In this equation, k is the number of features in the stimulus and the pattern, s(j) is the jth feature of the stimulus, p(j) is the jth feature in the pattern, sim(s(j),p(j)) is the similarity of feature s(j) to feature p(j), as specified by the similarity matrix, and w(j) is the weight of the jth feature in the pattern. The possibility that a stimulus may not contain all the features of a pattern (or vice versa) can be handled by including a null code within each *code* dictionary. The null code must not resemble any other code, i.e., the off-diagonal entries for the null row of the similarity matrix must be zero.

In some experiments patterns must differentiate between the appearance of a stimulus on an expected or an unexpected channel. This distinction is particularly important in studies of divided attention, where a person

may be told to react to the presence of a signal in the
right but not the left ear, or to a signal in the right
but not the left of the visual field. To allow for this
possibility, a pattern is further defined to have an
additional "feature," corresponding to the channel on
which the stimulus is expected. The importance of
stimulus location is specified by stating a channel
weight, c, varying from 0 to 1, where c=0 indicates that
the channel is irrelevant, and c=1 indicates that the
pattern is defined exclusively for one channel. The
strength of a match between a stimulus and a pattern is
then computed by the rule

$$
(2) \quad \text{Match between stimulus and pattern} = \begin{cases} \text{resemblance if the stimulus} \\ \text{is on the expected channel} \\ \\ (1-c)(\text{resemblance}) \text{ if the} \\ \text{stimulus is on an} \\ \text{unanticipated channel.} \end{cases}
$$

   The distinction between channels and features
deserves comment. Computationally, a channel is an array

variable that takes as its value a vector of features.
The pattern part of each production is similarly a vector
of features, and "perception" is the process of comparing
the value of a channel to the pattern part of a
production. This establishes a hierarchy of dimensions of
variation for stimuli. Stimuli may differ in their
features, and differ in the channel on which they appear.
The distinction is computationally important in the model,
because patterns are first matched to stimuli by computing
a weighted sum of matches based on corresponding features,
and the resulting value then multiplied by a weight
determined by the channel. An alternative scheme would be
to treat a channel as an additional feature of the
stimulus. Consider visual figures that varied in shape,
size, and color. Shape and size are determined by contour
and color is not. Should color be considered as a feature
to be added ... with shape and size, or should a simulation
be able to treat color in a completely different manner;
i.e. as a channel? We have chosen the alternative of
distinguishing between channels and features. The
implications of the other alternative have not been
studied.

   No claim is made that the similarity computation rule
is a theory of perception. (It would be of interest to

replace the rule with a psychologically more justifiable
one.) Using the similarity computation rule allows us to
proceed with the task of studying the post recognition
phenomena simulated by the Production Activation Model
itself. This point is developed in more detail in the
general discussion section.

The stimulus complex may contain information that
matches several patterns to varying degrees. Therefore
a "conflict resolution rule" is required to determine
which production is to have its associated action
executed. Conflict resolution is a general characteristic
of production systems (McDermott and Forgy, 1978). In the
Production Activation Model, the pattern part of each
production has associated with it a non-negative real
number, a, called its activation level. One, but only
one, of the factors involved in determining an activation
level is the extent to which the pattern matches some part
of the stimulus complex. The other factors are explained
below. The important point here is that the conflict
resolution rule is a procedure for comparing activation
levels.

As a computational device, the program's state is
computed for finite steps of time, called cycles. All

computations within a cycle take   ce functionally in
parallel. Within a cycle, at m(⟨ semantic, visual,
or acoustic pattern may be sele   'or execuation.
Therefore, patterns within a si   r~dality (or code) are
compared with one another to det ⬤f   nich one, if any
will be selected. The conditions ..r selection are

1. The pattern's activation level must be above a
preset threshold that is a characteristic of the pattern,
and

2. The activation level of the selected production
must exceed the activation level of any other pattern
stated in the same code by an amount, DELTA, that is a
parameter of the system.

The fact that the controlled system can respond to at
most one semantic, one visual, and one acoustic pattern
within a single cycle will be referred to as the
"bottleneck condition." The points at which productions
compete for action selection will be referred to as
bottleneck points. The Production Activation Model
contains three bottleneck points, one associated with each
of the codes.

A strong restriction of the model is that patterns
are stated in terms of the features on a single channel.
Thus it is not possible to define a pattern in terms of
concurrent features on two channels, i.e., in terms of a
multimodal stimulus complex. It is possible for the
system to react to multi-channel, multi-coded stimulus
complexes by recoding the components of these complexes to
internal stimuli in the semantic code, and then reacting
to the constructed semantic stimulus. The point is that
intervention by the controlled system is always required
to construct such an internal system. There is no way
that the system can react to multi-channel stimuli within
a single time cycle.

Any production system could be realized within the
controlled processing framework of this model. Since
unrestricted computing systems are equivalent to Turing
machines, it is generally held that they provide too much
power to be realistic psychological simulations. The
usual way to avoid the problem is to introduce the
restriction that patterns must not exceed a fixed length,
k^*, that could be looked upon as a limit on the capacity
of working memory. There are then only a finite number of
possible productions, corresponding to the finite number
of possibly discriminable stimuli, given that each

stimulus must consist of not more than k^* elements from a
finite code.

There is another restriction, not stateable in the
terminology of Turing machines but stateable by reference
to the Production Activation Model, that may be far more
important in limiting human capacity. It is the concept of
interruptability. Furthermore, this restriction interacts
with limitations in the size of working memory. Imagine
that the Production Activation Model, or some similar
device, contains a production system that is logically
sufficient to do an arbitrarily chosen calculation after a
minimum of n steps (n > 0). Suppose further that there is
some probability, q, that any arbitrarily chosen step may
fail to execute because of an interruption. That is, the
device is thought of as being embedded in an environment
in which high priority stimuli for productions outside of
the set in question appeared randomly on an external
channel. If such a signal appears, its processing takes
priority over the computaton currently being done. Let
P(n) be the probability of completing an n step
computation. Then

$$(3) \qquad P(n) = (1-q)^n$$

which becomes arbitrarily close to zero as n increases.

Finally, suppose that the size of a production pattern is
limited. The effect of this limitation would be to force
a large computation to be broken down into several steps,
thus increasing n. Clearly the computing power of the
model is limited by the size of the patterns that it can
recognize, and that limitation is exacerbated by the fact
that the system is interruptible by "irrelevant" stimuli.

The Automatic Processing System

The automatic processing system operates in a quite
different manner than the controlled system. It is best
understood by conceptualizing each production as a node in
a network that is similar to the semantic networks
described by Collins and Loftus (1975) and Anderson
(1976,1983, a,b). The connection between any two
productions, i and j, is stated as an association value,
a(i,j), that takes some value between 1 and -1. If
production i has activation level x(i,t) at time t, the
activation level of production j will be increased by the
amount a(i,j) * x(i,t) at time t+1. Thus information is
passed from production to production by spreading
activation, avoiding bottleneck points. All productions
transmit information about their activation level
simultaneously. A single production may send and receive

activation from several productions, including itself. If
the association link between two productions is negative
(a(i,j) < 0) the sending production is said to inhibit the
receiving production.

The two choice reaction time example can be extended
to illustrate the automatic information processing system.
Here it is useful to think of Stimuli 1 and 2 as
associated with Responses 1 and 2 as a result of
instructions, training and/or stimulus-response
compatibility. For instance, suppose that Stimulus 1 is a
right arrow ("->") and Stimulus 2 a left arrow ("<-"), and
that Responses 1 and 2 consist of the movement of a lever
to the right or left, respectively. Figure 2 shows the
network of associations that would be used to simulate
this situation. Three principles were used in
constructing the network (1). They are

1. All productions activate themselves, positively.
That is, 0 < a(i,i) < 1 for all i.

2. Any production whose action might produce the
precondition for a subsequent production is positively
associated with the subsequent production. Thus the
production recognizing Stimulus 1 primes the production

that would recognize the associated semantic signal. (To illustrate, a(v1,s1) > 0 in Figure 2.)

3. If the pattern parts of two productions represent logically exclusive interpretations of the stimulus, then the two productions inhibit each other. For this reason, a(v1,v2) < 0 and a(v2,v1) < 0 in the figure.

These rules were chosen because they have been found useful in a number of studies of self organizing systems. Note that Rule 3 is a logical analog to lateral inhibition, a phenomenon widely observed in the nervous system. The rules have been applied to the construction of semantic networks in all our simulation studies. In addition, only one value each is permitted for all self associations (a(i,i)), all positive associations, and all negative associations. The same values were used in all the studies reported here. The fact that reasonable results could be obtained without recalculation of parameters indicates that the model is robust.

Figure 2 here

Decay, Noise, and Refractoriness

If the automatic system operated exactly as described, activation would spread through the system without limit. To avoid this, a decay mechanism has been introduced. At the end of each time cycle, all activation levels are reduced by a fixed fraction d (0 < d < 1). (In all of our work, d has been set to .5) In addition, biological information processing systems are assumed to be subject to minor perturbations. These perturbations are modeled by adding a randomly distributed noise element to each production's activation level during each cycle. The noise element is distributed normally with expectation of zero and a standard deviation, e. The e parameter is thought of as a fixed characteristic of an individual at a given point in time. Except where noted, a constant value of e was used in all simulations. Similar decay and noise processes are required in virtually every associative network model of learning and cognition.

In production executing systems, the activation level of a production must be reduced once its action is taken. Otherwise the system will keep repeating its selection of the same production until new input is received (McDermott

and Forgy, 1978). In order to avoid this undesireable
outcome, a refractory process has been introduced.

When a production is selected for activation its
threshold is reset to a value halfway between its original
threshold and its current activation level. Subsequently,
the threshold decays toward its original level at a rate
determined by the decay parameter, d. This induces a
refractory period, during which time new productions can
be activated.

Selected Results

This section presents a brief description of selected
results that have been obtained with the simulation
program. Results have been chosen to illustrate
principles rather than to present a complete explanation
of the phenomena in question. Four experimental paradigms
will be considered. The first is the choice reaction time
paradigm. In choice reaction time studies the participant
knows that exactly one of a limited number of stimuli will
be presented on each trial. The task is to identify the
stimulus as rapidly as possible. This situation was
chosen for study because it represents a protypically
simple stimulus identification situation. Repetition
paradigms complicate the choice reaction paradigm by
introducing carryover effects from one trial to another.

(In "pure" choice reaction studies such effects are
statistically controlled by randomizing the sequence of
stimulus presentations.) Choice reaction time and
repetition paradigms deal with the identification of a
single stimulus, that can be treated as a unitary percept.
In both paradigms the participant knows where and when the
stimulus will appear; the only problem is to identify it.
By contrast, in dual task studies; two or more stimuli may
appear on a single trial. Typically the stimuli will
appear at different places in the participant's perceptual
field. The participant must split attention across the
different sources of input. Finally, in "Stroop"
paradigms two different stimuli are presented
simultaneously. The two stimuli may be associated with
incompatible responses. The participant's task is to
attend selectively to a preselected stimulus category,
while ignoring other stimuli.

Our simuluation studies have focussed on simple
versions of each of these four paradigms, in order to
illustrate the basic principles of the model. The
numerous complications that are possible within each
paradigm will not be dealt with here. Future papers that
explore the simulation of specific paradigms in greater
detail.

Choice Reaction Time Studies

Choice reaction time studies have produced a number of highly replicable phenomena, some of which have assumed the status of "laws" in psychology.

1. **Hick's Law**. The illustrative example dealt with a two choice reaction time study, i.e., on any trial, one of two stimuli appear. Choice reaction time studies may be extended to n-choice studies, by allowing the presentation of any one of n pre-defined stimuli on a given trial. As before, each stimulus is associated with a single *response*. If each of the n stimuli are equiprobable, average reaction time increases as a function of the logarithm of the *number of possible stimuli*. This finding is known as "Hick's Law."

An n-choice reaction time study was simulated by expanding the production systems and networks shown in Table 1 and Figure 1 to allow for 2, 4, or 8 stimuli and responses. The results are shown in Figure 3, where the number of time which plots the relation between number of choices (on a logarithmic scale) and the number of time cycles before a response. "Response" refers to the execution of a production whose action included an

external reponse. In this graph and all others each point was determined by averaging over roughly 1000 equivalent trials. Figure 3 also shows results of a study by Taylor (1982), using human subjects. Clearly the results with the simulation mimic the human results up to a ratio transformation (milliseconds per cycle).

Figure 3 here
----

2. **Speed-accuracy tradeoffs**. The relation between speed and accuracy of responding in choice situations has been the subject of considerable study. If a person speeds up his/her response in a particular choice situation the probability of an error increases. The relation between probability of correct response and time taken to respond is almost always a monotonically increasing, negatively accelerated function (Pachella, 1974). On the other hand, changes in either the *conditions of the task* or the *state of the individual* that produce slower reaction times almost always also increase the frequency of errors. These two effects will be refered to as the negative and positive speed-accuracy relations.

Both the negative and the positive relations can be

produced by the simulation, by manipulating different
parameters. Figure 4 shows the effect of manipulating the
DELTA parameter. Recall that this parameter determines the
extent of dominance that a production must have over its
competitors, before its associated action is taken.
Loosely, at high values of DELTA conflict resolution takes
longer, but is less likely to result from random
fluctuations in production activation levels. Thus
manipulating DELTA will produce a negative speed-accuracy
relation.

Figure 5 shows two different ways of producing
positive speed-accuracy relations. Figure 5A was produced
by holding the DELTA parameter constant and varying the
size of the noise parameter. Increasing the values of the
noise parameter both slowed responding and decreased
accuracy. This can be thought of as being an analog of
studies that compares responding across people of
differing information processing characteristics, e.g.
people of markedly different ages. Figure 5B shows a
positive speed-accuracy relation produced by holding the
DELTA and noise parameters constant, and varying the
parameter establishing the similarity between the two
visual stimuli. This is analogous to plotting data from
an experiment in which the stimuli to be identified vary

in distinctiveness.

Figures 4, 5 here

Stimulus repetition effects : In two choice CRT
experiments response time is a function both of the
choices available on the current trial and the relation
between the current and the previous trials. There are two
aspects to this relationship; the effect of the sequence
of stimuli presented and the effect of the
response-stimulus interval, i.e. the time between the
occurence of a response and the presentation of a new
stimulus. The two variables interact to produce fairly
wide variations in the response time. A subsequent paper
will present the simulation of inter-trial effects in
considerable detail. Here only the basic phenomena will
be presented.

A repetition is defined to be the presentation of
the same sequence on two or more successive trials. An
alternation is defined to be the presentation of
different stimuli on successive trials. While an
alternation could be defined for experiments involving any

number of stimuli, only the 2-choice experiment will be
considered here. Thus if the stimuli are arbitrarily
labeled 1 and 2 a repetitive sequence would be a sequences
of repeated presentations, as in 1..1..1..1, while
alternation would be represented by the sequence
1..2..1..2. An experiment by Kirby (1976) provides a good
illustration of the basic phenomena. Responses to
repetitive sequences were rapid at a response-stimulus
interval (RSI) of 50 msecs, while responses to alternating
sequences were rapid at an RSI of 2000 msec.

The production activation model accounts for the
short RSI repetition effects solely by the activation of
the automatic system. Assume that a correct response to
stimulus 1 has just been produced. By definition, the
activation levels of the productions associated with
identification of, and responding to, stimulus 1 will have
been higher than the activation levels of any competing
production. In fact, the only thing that keeps the
response from being repeated is the temporarily elevated
response threshold of the production producing the
response. If the same stimulus information is
re-presented, on the next trial, the automatic system will
already be biased toward production of the appropriate
response. However, the bias can be eliminated if the RSI

is sufficiently long so that the decay process reduces the
activation level of all productions to some low, near-zero
value.

Alternation effects are generally accounted for by
assuming that the decision maker has a more or less
conscious bias against expecting stimulus repetitions,
i.e. a crude version of the famous gambler's fallacy
(Kirby, 1976). This assumption can be modeled by assuming
that there exist productions in the controlled system that
bias the program toward alternation. To simulate such an
effect two additional productions were added to the
production system for the choice reaction simulation. The
first production had as its pattern the stimulus that
response 1 had occurred, the other pattern recognized that
response 2 had occurred. Upon recognition of response 1,
as a stimulus, the appropriate production placed an
internally generated stimulus that resembled stimulus 2 in
the visual working memory channel. In computing terms,
"resembled" simply means that the stimulus similarity
between the actual stimulus 2 and its internally generated
expectation was greater than zero, though less than 1, and
was also greater than the resemblance between the internal
stimulus and stimulus 1. Following the terminology used by
Posner (1978), this can be thought of as an internally

generated signal that primes the productions associated
with identifying stimulus 2. The production for
recognizing response 2 primed the productions associated
with recognizing stimulus 1 in a similar manner. Clearly
these productions, acting alone, would produce an
alternation effect, because the presence of the priming
signal in working memory would effectively lower the
thresholds of all productions associated with the expected
stimulus. However, the alternation mechanism can only be
effective if the RSI is long enough for the relevant
productions to be selected. This contrasts with the
repetition mechanism, which can only be effective at short
RSIs.

Figure 6 demonstrates response and alternation
effects at short response-stimulus intervals. Panel A
shows the effects in data produced by the
production-activation model. Sequence AAAA represents the
repetitive presentation of the same stimulus on four
successive trials, sequence ABAB represents presentation
of alternating stimuli. In the runs that produced this
data stimuli were presented immediately after the model
had made a response. There is a strong repetition effect
and no alternation effect. Panel B of Figure 6 is a
replotting of Kirby's (1976) data for his 50 msec RSI

condition. As in the model, there is a repetition effect
but not an alternation effect.

Figure 7 presents similar data for long RSI's. Panel
A shows the result of model runs that contained and
introducing a blank period (i.e. no stimulus present) of
six internal cycles between the response and the next
stimulus presentation. During this period the model
continued to redistribute activation, and to react to any
internally generated stimuli in the manner described
above. Panel B is a replotting of Kirby's data for the
2000 msec. RSI condition. In both the simulated and the
human data, repetition effects do no appear, but
alternation effects do.

In the human data, responses at the 2000 RSI
condition were more rapid, overall, than responses in the
50 msec RSI condition. This is not true in the model. No
attempt has been made to reproduce this effect, which may
be due to properties of the motor system rather than to
the interaction between expectation and stimulus
identification.

Figures 6 and 7 here

Splitting attention.   In choice reaction time studies the participant must identify a stimulus that appears in a known location. The next series of studies involve the detection of stimuli that appear randomly at different locations. An experiment by Kinchla (1980) provided the motivation for the simulation. Kinchla's observers had to attend to two small lights, at different locations in the visual field. In terms of the model, the two locations were treated as separate external visual channels. On each trial one or the other of the lights might flicker briefly. The observer's task was to indicate whether or not a flicker had occurred on either channel. Thus the dependent variable was the probability of a correct detection, rather than the latency of a response. The chief independent variable in the experiment was the priority that the observer was to assign to each channel. Priority was determined by instructions, and by points rewarded for a correct detection.

    The simulation for this task was closely related to the CRT simulation. Information was presented over the two

visual external channels. Initially both channels had null signals (S1) placed on them. These signals corresponded to the lights. Then, on experimental trials, a target (S2) signal was placed on one of the channels briefly. Catch trials were also included, in which no target signal was presented. (Naturally, Kinchla also used catch trials.) The target signal was then replaced with the S1 signal.

    To permit false alarms, which do occur in this type of study, the resemblance between S1 and S2 stimuli was set at .50.

    The production system used is presented in Table 1. It contains two productions associated with each visual channel. The two productions can be thought of as miniature "choice reaction" productions within each channel. One recognized stimulus S1, and the other recognized stimulus S2.   A trial on which a signal was presented was judged correct if either of the productions for recognizing S2 was selected for execution. On catch trials, the firing of one of these production was recorded as a false alarm.

Table 1 Here

Figure 8 shows the semantic activation network associated with the production system. All productions were assigned to the visual code. Since all productions were at the same level (detection of an elementary visual display) each production positively primed itself, and inhibited all other productons.

The notion of priming was used to simulate the effect of instructions, in a manner similar to that developed for the study of alternation effects. It was assumed that, given appropriate instructions, a person could generate an internal "priming" signal that would have the effect of lowering the response thresholds of all productions associated with a particuar channel. Thus the threshold value for each channel served as the primary dependent variable. Thresholds varied between 1 and 0, complementarily. That is, if threshold x was assigned to the productions of channel 1, threshold 1-x was assigned to the productions of channel 2.

Figure 8 Here

Date from this sort of study is usually represented as a "performance operating characteristic" (POC), in which the accuracy of detection of targets on one channel is plotted against the accuracy of detection of targets on the other channel. In Kinchla's (1980) study the POC was linear, providing that performance was plotted in terms of "hit rate", the percent of targets correctly detected. For the POC to be meaningful, though, false alarms should occur on only a small percentage of the trials. (It is important to have some false alarms, to ensure that the subject has not set an excessively high criterion for target detections (Pachella, 1974).)

Figure 9 presents the POC obtained by the simulation. It is clearly linear. The error rate for catch trials was about 5% overall, with somewhat more false alarms occurring when high priority was given to one or the other of the channels. Whether or not the last effect is characteristic of human responding is unknown. Kinchla's data is also shown for comparison. Note that in Kinchla's

study subjects never completely ignored the less relevant
stimulus, so the POC does not cover the extreme points
that could be simulated.

Figure 9 Here

Stroop studies . The last simulation to be reported deals
with the Stroop task, a situation considerably more
complicated than the other paradigms described here. In
Stroop situations signals are presented simultaneously on
two separate channels. The participant is instructed to
make an identifying response to the stimulus on one
channel, while ignoring the other. The two stimuli will be
referred to as the relevant and irrelevant stimuli,
respectively. There are three basic experimental
conditions. In the neutral condition the relevant and
irrelevant stimuli are not associated with either common
or mutually exclusive responses. In the conflict
condition the relevant and irrelevant stimuli have strong,
and mutually contradictory, associations with the possible
responses. In the facilitating condition the relevant and
irrelevant stimuli are both highly overlearned cues for
the same response.

There are a variety of Stroop situations, each of
them producing a somewhat different pattern of responses.
This report deals only with a simulation of a standard
Stroop task, based on Stroop's (1935) original experiment.
(A much more detailed study of Stroop tasks will be
reported subsequently.)   In Stroop's experiment
participants had to name the color of ink in which words
were printed. The words were themselves color names; e.g.
the word GREEN printed in red ink. This, obviously, is a
conflict condition. In a facilitation condition GREEN
would be printed in green ink, while in a neutral
condition either a color patch or a non-color word (e.g.
DOOR) would be presented in colored ink. In general, if
people are asked to name the ink color there is a marked
slowing of identification responses in the conflict
condition, and a negligible increase of speed in the
facilitation condition (Dyer, 1973). Reading the word
takes half as much time as color naming, and is relatively
uninfluenced by ink color.

The simulation of performance in the Stroop task was
based on the model proposed by Morton (1969). Morton used
a somewhat specialized vocabulary, so his ideas will be
restated using the vocabulary of the Production Activation

Model. For those who are familiar with Morton's notation,
his logogen corresponds roughly to a node in the semantic
network. In Morton's model there is no element
corresponding directly to our "production," but the idea
of a production's being activated seems in the spirit of
his notion of the activation of logogens.

Stimuli are assumed to appear on two visual channels,
as in the case of the simulation of split attention. The
actions subsequent to recognizing visual stimuli, though,
are much more complex than in the case of the detection
experiment. Table 2 shows the production system for a
simple Stroop task in which the words RED and GREEN might
be presented in red and green ink. It consists of three
parts. The visual productions are somewhat analogous to
those in the detection study; each channel has associated
with it productions for recognizing either the words or
colors for "red" or "green." The action parts of these
productions each generate an appropriate semantic
stimulus, indicating to the system that a "red" or "green"
stimulus has occurred. The semantic stimulus, in turn, is
recognized by semantic productions that generate an
internal auditory code indicating to the system that an
external verbal response is required. When the auditory
internal signal is recognized the appropriate verbal

response is made. (Naturally, the computer does not
actually make this response. It is simply recorded as
being executed.)

Table 2 Here

The two part translation from visual to semantic
codes, and from semantic to verbal responding, may seem
clumsy. It was included to extend the simulation to
situations in which external instances of internal
concepts must be recognized, and the concepts themselves
must be used to guide the selection of "response codes",
usually for auditory responses. See Morton (1969) for
further justification of the basic theory.

The visual productions contain two features in their
pattern part; one referring to the external stimulus and
one referring to the instructions. Instructions are
presumed to act as internally generated goal stimuli that
prime particular productions, in the manner described
earlier for the split attention and stimulus alternation
studies. Thus it is possible for a stimulus to fully match
a visual production only if the stimulus is relevant to

the instructions given. However irrelevant stimuli may
partially match their associated productions.

Figure 10 displays the semantic network associated
with the productions. Several assumptions from Morton's
model are incorporated in the figure. One is that, as a
result of overlearning, the forms of words (and hence
their associated productions) will have become associated
directly with the acoustic codes for words. By contrast,
color stimuli, and their associated productions, are
assumed to be associated with the semantic codes for
colors, rather than the name codes. However name codes and
semantic codes for colors are both assumed to be
associates of each other. The effect of these assumptions
is that in a Stroop conflict situation the controlled
production system and the automatic semantic network
system partially work against each other. Irrelevant word
stimuli *should be more effective interferers than*
irrelevant color stimuli because the word stimuli have a
rapid path to the (incorrect) response, via the automatic
processing system.

Figure 10 here

Figure 11 depicts the results of a simulation of the
six possible Stroop conditions. Most of the Stroop
findings are replicated. The color naming-conflict
condition produces by far the slowest responding.
Facilitation effects are relatively small for color
naming, and non-existent for word naming. There are two
exceptions to the normal Stroop finding. One is that the
"word reading" conditions are not as much faster than
"color naming" as they should be. The second
contradiction to human data is that conflict is found in
the word reading condition. That is, it takes longer for
the simulation to "read" the word RED printed in green ink
than it takes to read RED printed in a neutral color. A
detailed analysis of the simulation indicates that these
effects could be markedly reduced by increasing the power
of the forward association between word recognition and
auditory pattern recognition. This would involve
manipulating a parameter that has been a constant in all
other studies reported here. While the manipulation seems
a reasonable one, the data in Figure 10 are reported for
consistency, and because our aim is to show that the same

simulation can handle the gross phenomena from a variety
of paradigms, rather than to demonstrate how the model can
be tuned to mimic the data from a single situation in
great detail.

Figure 11 here

Related Theoretical Studies

This section will compare the Production Activation
Model to three other theoretical proposals; the ACT* model
developed by Anderson (1983a,b), the cascade processing
model developed by McClelland (1979; Rumelhart and
McClelland, 1982), and Eich's CHARM model of memory
storage and retrieval. These models should not be
considered as contending explanations of the same
phenomena analysed using the Production Activation Model,
nor do they contend with each other. The different models
represent complementary and compatible approaches to
distinct, but related, cognitive behaviors.

The ACT* model

Anderson's ACT* model of memory and problem solving
resembles the Production Activation Model in several ways.
Both are systems for interpreting productions in order to
determine actions. Both represent information in long term
memory using semantic networks. ACT* contains three types
of internal stimulus codes; codes for spatial information,
temporal order, and for abstract propositional
information. The three codes are loosely analagous to the
visual auditory, and semantic codes of the Production
Activation Model.

The most striking discrepancy between the two models
is in the complexity of the patterns that serve as cues to
production activation. In ACT* production patterns may be
complex conditional statements. For example, Anderson
(1983b, pg. 144) presents an eighteen term conjunction as
the pattern part of a production for recognizing the word
EACH. Several of the terms in this conjunction are
themselves conjunctions. Such patterns are clearly much
more complex than the patterns dealt with in the present
studies of the Production Activation Model. However,
complexity alone does not represent a basic distinction.
It would be possible to represent Anderson's eighteen term

conjunction within the Production Activation framework,
and there is every reason to believe that people do have
the ability to recognize complex situations. A more basic
difference between the models is in the techniques that
are used to match stimuli to patterns.

In ACT* patterns are matched to stimuli using a
computational method known as data-flow pattern matching
(Forgy, 1979). Explaining the details of such a mapping
would be beyond the scope of this paper. The important
point is that a commitment to the data flow matching does
not restrict the logical complexity of the description of
the class of stimuli that activate a given production (1).
By contrast, the Production Activation Model uses a linear
discrimination procedure to determine the relative
strengths of matches between the current stimulus and the
patterns stored in long term memory. There are
interesting, relatively simple classification rules that
cannot be realized by a linear discrimination procedure
(Minsky and Papert, 1968). For instance, the
biconditional rule (An item is a member of class X if it
has both features a and b or if it has neither feature)
cannot be realized by such a procedure, although it could
be realized by data flow pattern matching. Therefore more
powerful productions can be stated within the ACT*

framework than within the Production Activation framework.
Whether this should be regarded as a strength or weakness
is a matter of debate. Linear systems can be justified by
reference to simple neural models. However linear systems
are probably not powerful enough to model all human
pattern recognition. Does such modeling require a system
of the complexity of data flow networks? The answer to
this question is not known.

A similar remark can be made about the use of codes
to define the pattern parts of productions. The Production
Activation Model contains the strong assumption that
patterns are defined only within a code type. It is for
this reason that productions compete for activation only
within classes of productions; auditory productons with
themselves but not with visual productions, and so forth.
Although ACT* has three code types, there is apparently no
requirement that the pattern of a production be restricted
using only one code type. ACT* could recognize a
cross-modal stimulus, such as a barking dog, by activating
a single production. Thus ACT* assumes a primitive ability
for cross-modal integration, while the production
activation model requires a sub-model of the integration
process.

In the Pattern Activation Model a stimulus is defined
as an ordered set of features on a particular channel.
ACT* does not have a similar concept of channels within a
code type. Thus ACT* has the capability of reacting
directly (i.e. by the matching of a single production) to
a stimulus defined by the totality of features present in
the external world, while the Production Activation Model
must build up an internal representation of that stimulus,
by recognizing the presence of patterns on different
stimulus channels, and using these recognitions to piece
together an internally generated coding of the total
stimulus complex.

The distinction between production activations based
on restricted or unrestricted pattern recognition is a
serious theoretical issue. In the ACT* framework a cross
modal stimulus, or a cross-channel stimulus within a
particular mode of presentation, could activate a
production directly. Therefore it could initiate automatic
data processing within the semantic network. While
something resembling cross modal (or cross channel)
automatic data processing could occur in the Production
Activation model, through partial arousal of patterns, the
amount of automatic information processing would be
limited. This raises an important empirical question. Is

it possible to automate responding to stimuli defined in
terms of more than one internal code? Or defined in terms
of the combination of information on more than one
channel?

ACT* and the Production Activation model also differ
markedly in their treatment of reaction times. In the
Production Activation system response times are derived
from the dynamics of the development of activation levels
associated with the various productions. In ACT* reaction
times for memory retrieval are computed as a function of
properties of the distribution of activation in a stable
semantic network, after the dynamic phase of a response to
a stimulus has been completed. The difference in emphasis
is appropriate, considering the difference in the type of
process being studied. The Production Activation Model
studies have been focussed on situations requiring rapid
responses to easily recognized stimuli. The response
times being modeled are seldom more than a second. The
retrieval processes studied in ACT* simulations (Anderson,
1983a) may take several seconds. It seems reasonable to
regard them as the result of a search process that
examines stable activity of a semantic network, instead of
the being an offshoot of the initial dynamic response of
the system to stimulus presentation. Anderson does not

offer a model of the search process itself.

These differences between the models can be resolved
by regarding ACT° and the Production Activation Model as
complementary models that deal with different phenomena,
in a compatible manner. ACT° models have been proposed
for the retrieval of information from memory, and for
complex cognitive acts, such as solving problems in plane
geometry. An analysis of such actions by models that
began with phenomena at the level of physical stimulus
recognition would quickly become hopelessly complicated.
The Production Activation Model has been designed to deal
with these precisely more molecular situations (2).

We conjecture that ACT° could be recreated as a set
of productions to be executed by the Production Activation
Model. Data flow networks, which are treated as
primitives by ACT°, would be treated as decision processes
to be modeled. The complex patterns recognised directly
by ACT° would then be "named", as a response, by the
Production System model. Similarly, cross modal
integration could be simulated using the Production
Activation model, and the output of the simulation used as
an "elementary" stimulus feature in the ACT° framework.
The resulting simulation program would almost certainly be

a clumsy computational device, so one would hardly propose
developing a working model for use in extensive
simulations. The development of one or two simulations to
test the extent of complementarity would be interesting
exercise in theory development.

There are some minor discrepancies between the models
that require brief comment. ACT° and the Production
Activation model differ in the details of the algorithms
that are used to execute various actions; e.g. to compute
the spread of activation levels through the semantic
network. Given the current state of knowledge concerning
the simulation of cognitive action, the use of slightly
different algorithms is a minor discrepancy. At some
point, though, a theoretical justification should be
offered for all algorithms used in a simulation. This
point will be discussed again, in connection with
McClelland's cascade model of information processing.

Finally, ACT° is programmed in LISP, while the
Production Activation Model has been programmed in PASCAL.
This discrepancy is purely a technical one, and has no
psychological significance whatsoever.

Cascade models of pattern recognition

In the Production Activation model processes associated with the selection of a production at one stage of information processing can begin before the completion of processing at prior stages. For example, if a visual stimulus is presented the activation of relevant visual patterns can feed activation forward to semantic patterns, before any production breaks through the visual bottleneck point to be executed. This contrasts with serial models of information processing, which require the completion of analyses at one stage before a subsequent stage can begin. McClelland (1979) coined the term "Cascade Processing" to refer to models which have this characteristic. He presented a detailed description of a particular class of cascade models. Subsequently, Rumelhart and McClelland (1982) used the cascade model as the basis for a simulation of visual word recognition.

In McClelland's cascade model, concepts in memory are represented by nodes that vary in their level of activation. The nodes in a network are divided into subsets ("levels"). Connections between nodes on the same level are predominately inhibitory, while connections between nodes in different levels are generally

facilatory. This is exactly the scheme used to connect the nodes representing "semantic", "visual", and "auditory" patterns in the Production Activation Model (3). Therefore the networks used in the Production Activation Model could be thought of as specific examples of the cascade networks that McClelland describes.

McClelland's models have been applied to the simulation of word recognition of words (Rumelhart and McClelland, 1982; McClelland and Rumelhart, 1981). In these studies the process of recognition begins with the identification of very simple visual features, such as horizontal or vertical line segments, and word and letter recognition is derived. In the Production Activation studies stimulus recognition at the word level is asserted as a primitive process. It would be possible to regard McClelland's work as an analysis of how activation is spread from level to level, and the Production Activation studies as analyses of the consequence of that spread. The same view could be taken of the relation between McClelland's work and Anderson's.

McClelland's models deal only with pattern recognition and memory arousal. In the terminology of the Production Activation model, cascade processing is a

mechanism that could be used to produce automatic responding. *The cascade model does not deal with mechanisms for controlled responding.*

There is one conflict between McClelland's models and the models explored within the Production Activation framework. In McClelland's model the spread of activation from node to node is controlled by an algorithm that has *the desirable property of limiting the maximum activation level of any node to a pre-established asymptotic value.* The algorithm used in the Production Activation model could, in theory, produce an arbitrarily high level of activation at any one node. While this situation has not been encountered in practice, its possibility is bothersome because the potential for a limitless *activation level seems unrealistic. A few attempts have been made to reproduce the simulations reported here,* using McClelland's algorithm for spreading activation through the network. In none of the cases studied did the model's behavior resemble human data. Interestingly, Anderson (1983b) reports similar results when McClelland's algorithm was incorporated into ACT*. Given the success of the algorithm in simulating the data from stimulus recognition situations, its failure in these simulations is puzzling. Solving this puzzle would be a useful study

in the mathematical analysis of psychological theories.

The CHARM Model

Eich (1982) has developed a "Composite Holographic Association and Retrieval Model", hence the acronym. In the CHARM model a stimulus is a vector of feature strengths. Associations between two stimuli are established by storing the convolution of the two stimulus vectors. Storing the convolution of stimulus vectors produces an information storage system similar to the "memory" displayed by a holograph, which stores the convolution of the light spectra falling on it from different sources. Eich has shown that the information contained in the convolution may be combined with information in a probe stimulus (i.e. a recall cue) to reconstruct the original stimulus.

Eich's model does not contain a production system stage, which is not surprising since the model has been presented solely as a model of memory. CHARM is discussed here because it could be incorporated into the Production Activation Model as an explanation of the pattern recognition process that is a primitive in the Production Activation model. In the model specified here stimuli are

recognized by a computing the similarity between the
stimulus complex and the pattern recognition part of a
production. Similarity is defined to be a weighted sum of
the resemblance between features of the stimulus and
features of the pattern. How the pattern came to be in
long term memory is a problem outside of the scope of the
model. Eich's proposal provides motivation for an
alternative model, in which each pattern is a record of
the convolutions of the stimulus-stimulus sequences that
the system has experienced. If Eich's model were to be
incorporated as the pattern recognition process for
production activation, this would at once provide an
elementary learning mechanism and offer a mechanism for
pattern recognition that could be justified by appeal to
supporting psychological studies, rather than to the
assertion that it produces reasonable results.

   While the idea of incorporating both the cascade
processing and CHARM models within the Production
Activation framework is a conceptually appealing one, the
technical problems of incorporation should not be
underestimated. Both the CHARM and cascade processing
models require extensive numerical computation. The
computational burdens of the Production Activation model
itself are not trivial (4). The computational problem may

not be solvable without access to the very large, high
speed array processing computers that are just now
becoming available. The explorations of this class of
psychological models may be well be limited by the
computing power available to the researchers.

Summary

   The Production Activation Model incorporates two
ideas; the use of a semantic network to activate concepts
held in long term memory and the use of production systems
to manipulate information in working memory. The semantic
network is seen as controlling automatic, highly
overlearned responding. The working memory system
determines slower, controlled responding based on
decisions made about the current stimulus complex. The
model has been used to construct simulations of human
behavior in a variety of paradigms in which people must
choose quickly between possible interpretations of the
stimulus, monitor two stimulus channels for relevant
information, or respond to one channel while ignoring
another. The success of the simulations indicates that
the model has considerable breadth of applicaton.

   The logical relation between the Production

Activation Model and other models of memory and attention
was considered. The Production Activation Model deals
with processes more molecular than those described in
Anderson's ACT*. As a conjecture, we believe that the
principle features of ACT* could be reprogrammed as a
special case of the Production Activation model. In
turn, McClelland's models of cascade processing and Eich's
CHARM model of memory can be thought of as models for
processes that are treated as primitive actions in the
Production Activation Model.

Mathematical models of psychological phenomena have
been criticized for being elaborate models of highly
specialized laboratory paradigms. The work reported here
partially refutes such a claim. What has been shown is
that there is considerable compatability between the ideas
used in the analysis of such disparate situations as two
choice reaction time experiments and the analysis of
protocols taken from master chess players. A great deal of
further theoretical work remains to be done to explore
various areas of compatability and incompatability. Some
aspects of the further work will require extensive
computing facilities in order to conduct the necessary
simulations. Nevertheless, we are encouraged at the
prospects for development of psychological theories that

are at once broad and precise.

References

Anderson, J. R. (1976). _Language, memory, and thought._ Hillsdale, NJ: L. Erlbaum Assoc.

Anderson, J. R. (1983a). A spreading activation theory of memory. _J. Verbal learning and verbal behavior, 22,_ 261-295.

Anderson, J. R. (1983b). _The architecture of cognition._ Cambridge: Harvard U. Press.

Baddeley, A. D. (1976). _The psychology of memory._ New York: Basic Books.

Collins, A. M. and Loftus, E. F. (1975). A spreading activation theory of semantic processing. _Psych. Review, 82,_ 407-428.

Dyer, F. H. (1973). The stroop phenomenon and its use in the study of perceptual, cognitive and response processes. _Memory and Cognition. 1,_ 106-120.

Eich, J. M. (1982). A composite holographic associative

recall model. _Psych. Review, 89,_ 627-661.

Forgy, C. L. (1979). On the effect implementation of production systems. Ph.D. dissertation, computer science dept., Carnegie-Mellon University.

Hunt, E. B. (1967). A model of information acquisition and organization. Proceedings of symposium on human information processing. In F. Klix (Ed.), XVIII International congress of psychology, Moscow, 1966. Berlin: Akademie-Verlag Gmblt, 87-97.

Kinchla, R. A. (1980). The measurement of attention. In R. S. Nickerson (Ed.) _Attention and performance VIII._ Hillsdale, NJ: Erlbaum Assoc.

Kirby, N. H. (1976). Sequintial effects in two-choice fraction time: Automatic facilitation or subjective expectancy? _J. of Exp. Psych.: Human perception and performance, 2,_ (4), 567-577.

Kosslyn, S. L. (1980). _Image and mind._ Cambridge: Harvard U. Press.

McClelland, J. L. (1979). On the time relations of mental

processes: An examination of systems of processes in

cascade. Psychol. Review. 80. 287-330.

McClelland, J. L. and Rumelhart, D. E. (1981). An

interactive model of context effects in letter

perception: pt. I. An account of basic findings.

Psych. Review. 88. 375-407.

McDermott, J. and Forgy, C. (1978). Production system

conflict resolution strategies. In D. A. Waterman and

F. Hayes-Roth (Eds.), Pattern directed inference

systems. New York: Academic Press.

Minsky, M. and Papert, S. (1968). Perception. Cambridge:

MIT Press.

Morton, J. (1969). Categories of interference: Verbal

mediation and conflict in card sorting. Br. J. of

Psychol.. 60. (3), 329-346.

Newell, A. (1973). Production systems: Models of control

structures. In W. G. Chase (Ed.) Visual information

processing. New York: Academic Press.

Newell, A. and Simon, H. (1972). Human problem solving.

Englewood Cliffs, NJ: Prentice Hall, Inc.

Pachella, R. G. (1974). The interpretation of reaction

time in information processing research. In B. H.

Kantowitz (Ed.) Human information processing:

Tutorials in performance and cognition. Hillsdale, NJ:

L. Erlbaum Assoc.

Posner, M. I. (1978). Chronometric explorations of mind.

Hillsdale, NJ: L. Erlbaum Assoc.

Rumelhart, D. E. and McClelland, J. L. (1982). An

interactive activation model of context effects in

letter perception: pt.2. The contextual enhancement

effect and some tests and extensions of the model.

Psych. Review. 89. 60-94.

Schneider, W. and Shiffrin, R. M. (1977). Controlled and

automatic human information processing: I. detection,

search and attention. Psychol. Review. 84. 1-66.

Stroop, J. R. (1935). Studies of interference in serial

verbal reactions. J. Expert'l Psychol.. 18. 643-662.

Taylor, J. L. (1982). The effects of benzodiazepines on

cognition and performance. Ph.D dissertation,

Psychology Department. University of Washington.

Treisman A. and Gelade, G. (1980). A feature integration

theory of attention. Cognitive psychology, 12,

97-136.

Author's Notes

1. Anderson's data flow networks contain nodes that can
compute logical functions of their inputs. See, for
instance, the example provided in Anderson, 1983b, pg.
139-145. Given this much computing power, it is not
clear that there are any limits to data flow networks
other than those that apply to Turing machines. In
more psychological terms, ACT* provides no principles
to limit the complexity of patterns that can be
recognized directly.

2. Anderson (1983b) has sketched ways in which ACT*
might be applied to pattern recognition and attention
and performance paradigms similar to those studied

here. His detailed studies, however, have been of much
more complex situations involving learning, problem
solving, and long term memory retrieval (Anderson,
1983a,b).

3. The organization of nodes in the Production System
Model was based on earlier work by Hunt (1967).
McClelland's (1979) detailed analysis of cascade models
represents the parallel evolution of a similar idea.
The general concept of lateral inhibition coupled with
positive feed forward circuits also appears in a number
of other proposals for mathematical models of
biological information processing.

4. The simulations reported in this article required
several hours of central processor time on a VAX 780
computing system. This does not include time for
program development or for exploratory computations.

Figure Captions

Figure 1. A schematic of the architecture of the
production-activation model.

Figure 2. The semantic network used to connect
productions in simulating a two-choice reaction time
study. V[x] is the rule "if visual stimulus x is
recognized, create semantic stimulus S[x]." S[x] is the
rule "if semantic stimulus x is present make response x."

Figure 3. A simulation of Hick's law. Reaction time
increases logarithmically with the number of alternatives
in the model. Human data from Taylor's (1982) study shows
a similar relation.

Figure 4. Reaction time and accuracy are both increased
by increasing the DELTA paramenter. This mimics the
negative speed-accuracy relation.

Figure 5. Reaction time increases and accuracy decreases
if noise is added to the system internally (Figure 5a) or
if the similarity between stimuli is increased (Figure
5b).

Figure 6. Responding to repetitive stimuli (AAAA) of
alternating sequences (ABAB) as a function of the position
of the stimulus in a sequence. Panel A shows data from
the model. Panel B shows data from Kirby's (1976) study.
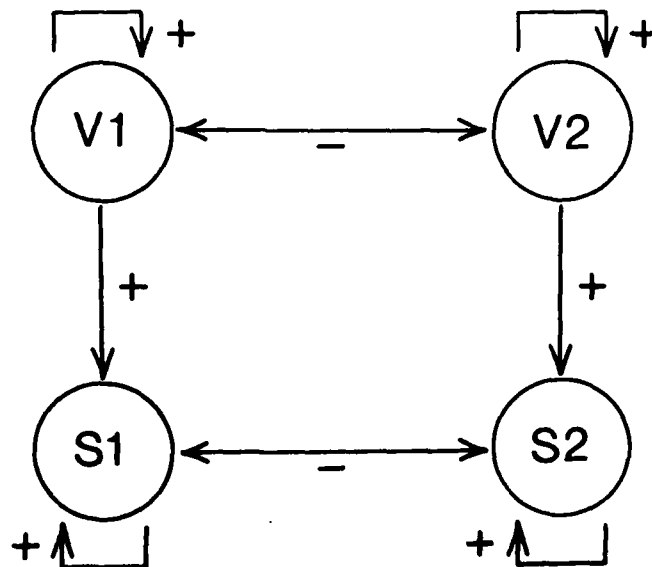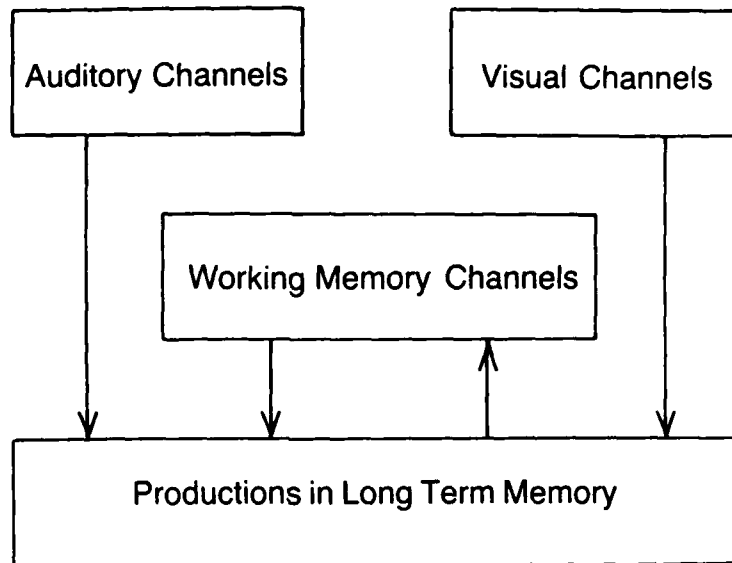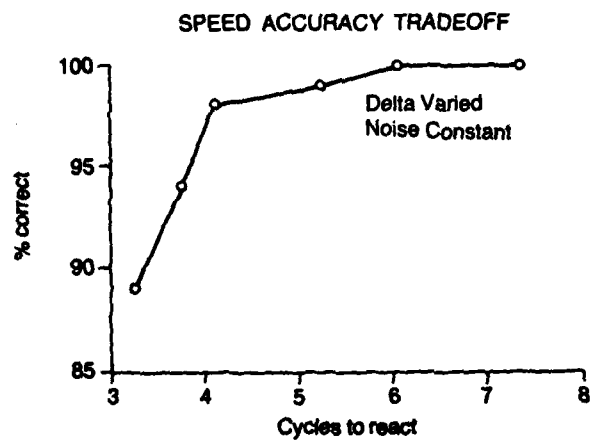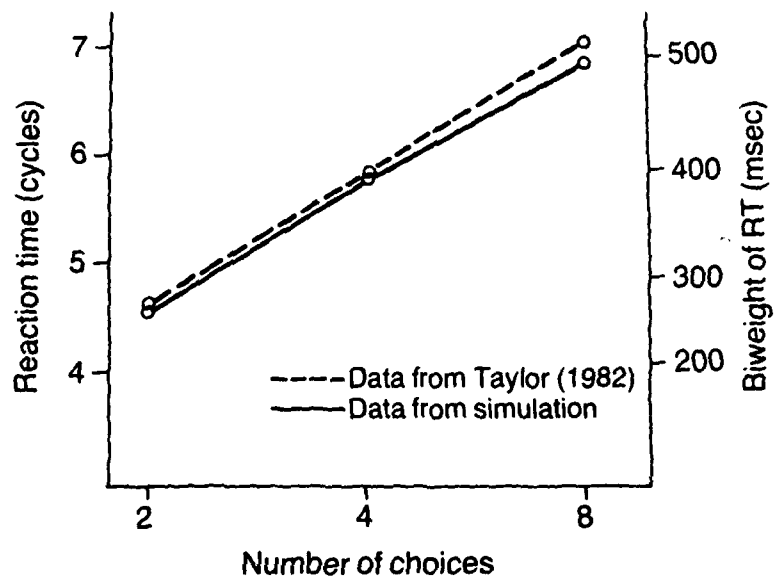Data is shown for short RSI conditions.

Figure 7. Responding to repetitive stimuli (AAAA) or
alternating sequences (ABAB) as a function of the position
of the stimulus in a sequence. Panel A shows data from
the model. Panel B shows data from Kirby's (1976) study.
Data is shown for long RSI conditions.

Figure 8. The semantic activation network used to
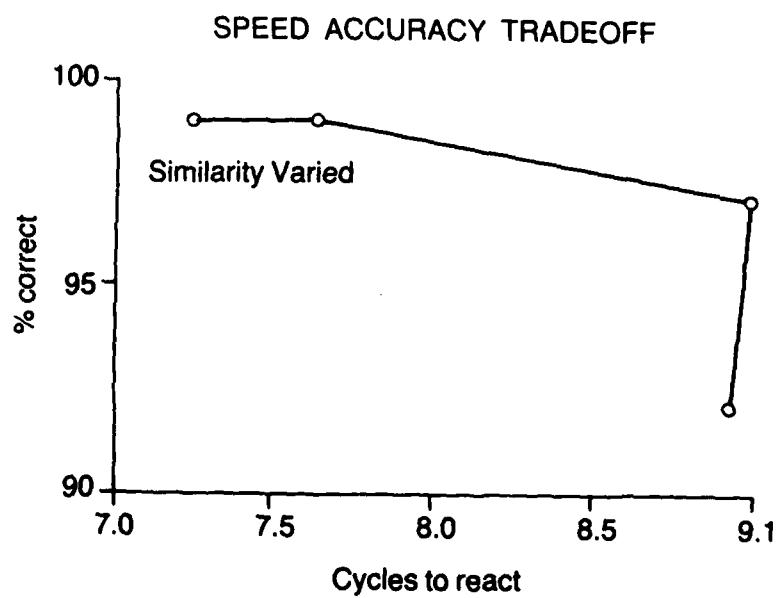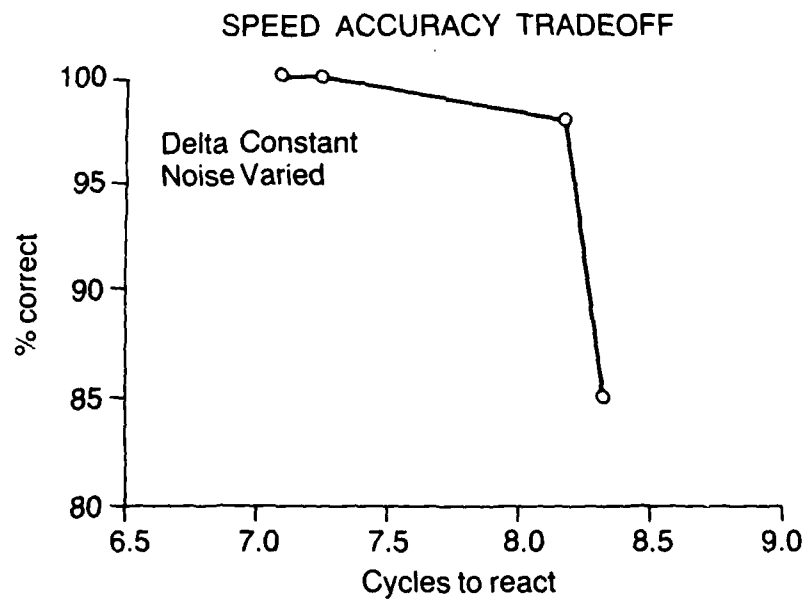simulate the simultaneous monitoring visual channels.

Figure 9. The performance operating characteristic for
monitoring two channels. Squares are data produced by the
model. Triangles are data replotted from Kinchla's (1980)
study.

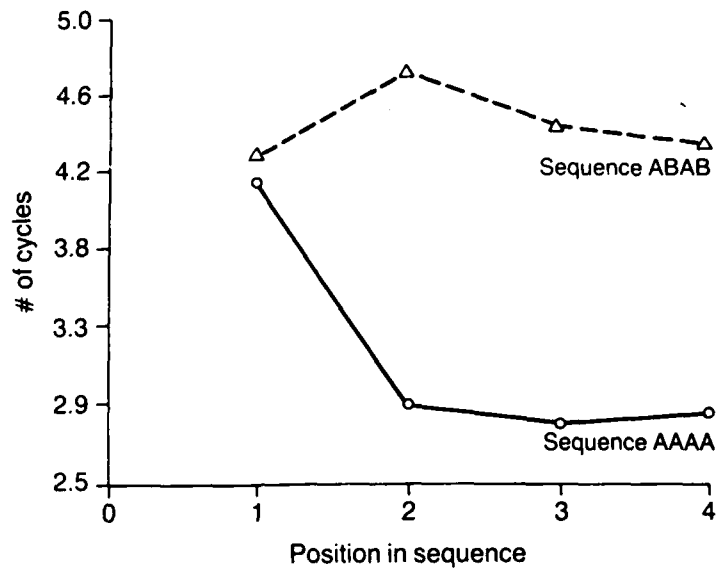Figure 10. The semantic network used to simulate stroop
phenomena.

Figure 11. Time to react as a function of conditions in
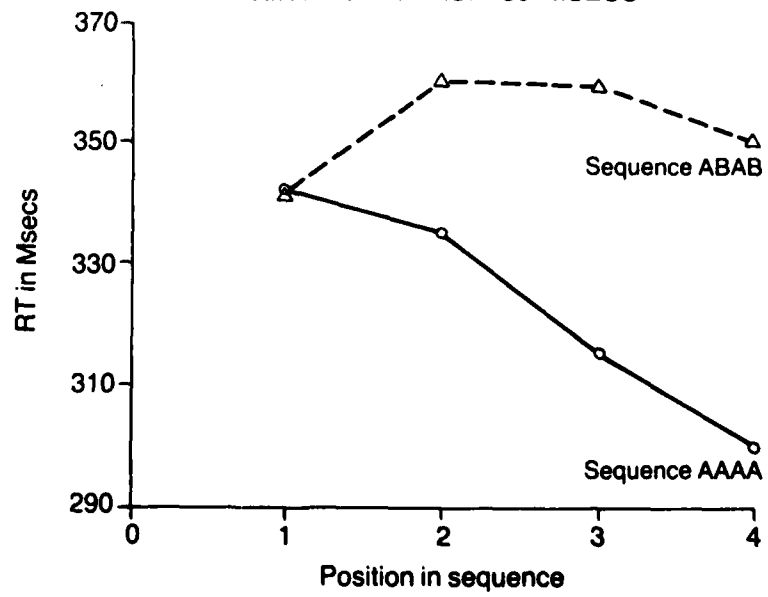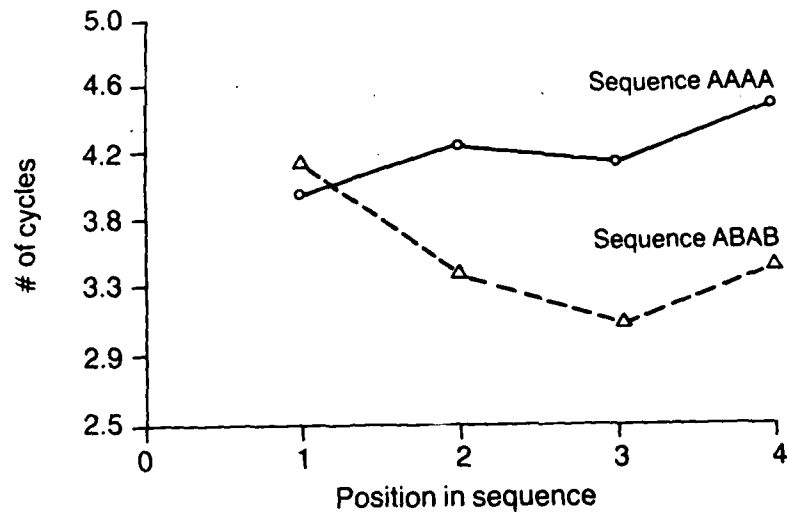standard stroop paradigm.

Auditory Channels

Visual Channels

Working Memory Channels

Productions in Long Term Memory

V1 — — → V2 $-$

V1 → S1 $+$

V2 → S2 $+$

S1 — — → S2 $-$

Reaction time (cycles) / Biweight of RT (msec) vs Number of choices

- - - - Data from Taylor (1982)
——— Data from simulation

SPEED ACCURACY TRADEOFF

Delta Varied
Noise Constant

% correct vs Cycles to react

SPEED ACCURACY TRADEOFF

Delta Constant
Noise Varied

% correct (y-axis: 80, 85, 90, 95, 100)

Cycles to react (x-axis: 6.5, 7.0, 7.5, 8.0, 8.5, 9.0)

SPEED ACCURACY TRADEOFF

Similarity Varied

% correct (y-axis: 90, 95, 100)

Cycles to react (x-axis: 7.0, 7.5, 8.0, 8.5, 9.1)

MODEL RSI = 0

Sequence ABAB

Sequence AAAA

KIRBY DATA RSI = 50 MSECS

Sequence ABAB

Sequence AAAA

MODEL RSI = 0

Sequence AAAA

Sequence ABAB

# of cycles

Position in sequence

KIRBY DATA RSI = 2000 MSECS

Sequence AAAA

Sequence ABAB

RT in Msecs

Position in sequence
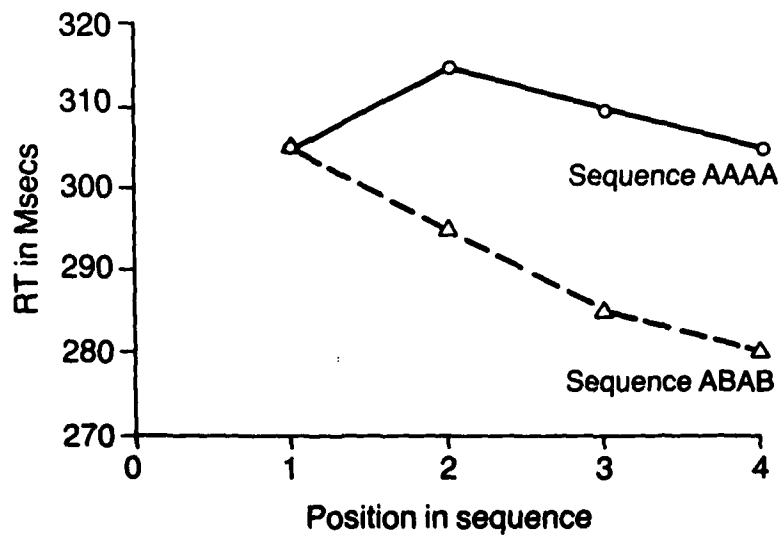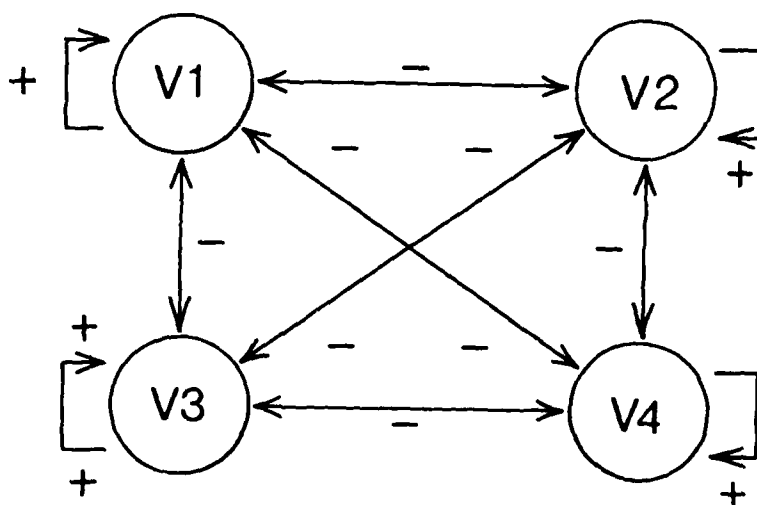
POC FOR DUAL VISUAL DETECTION

Form
red

Color
red

Color
green

Form
green

Semantic
red

Semantic
green

Auditory
red

Auditory
green

15

14

13

12

11

Reaction times (cycles)

□ name word
o name color

1
Facilitation

2
Neutral

3
Conflict

Stroop study
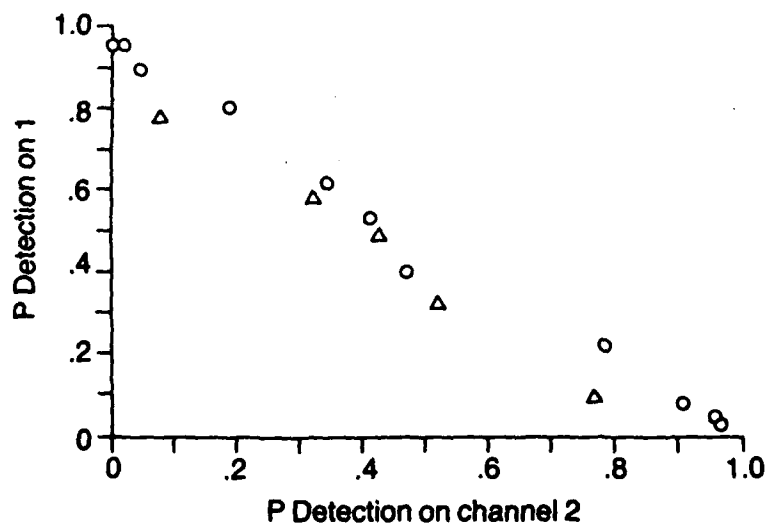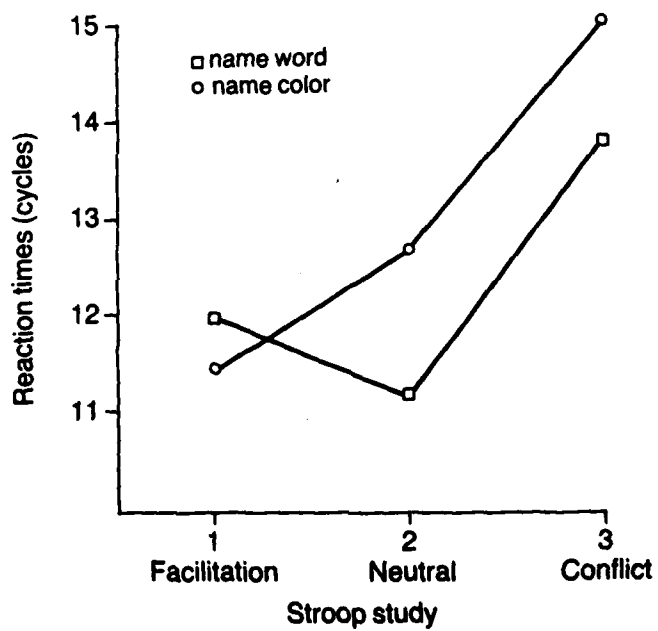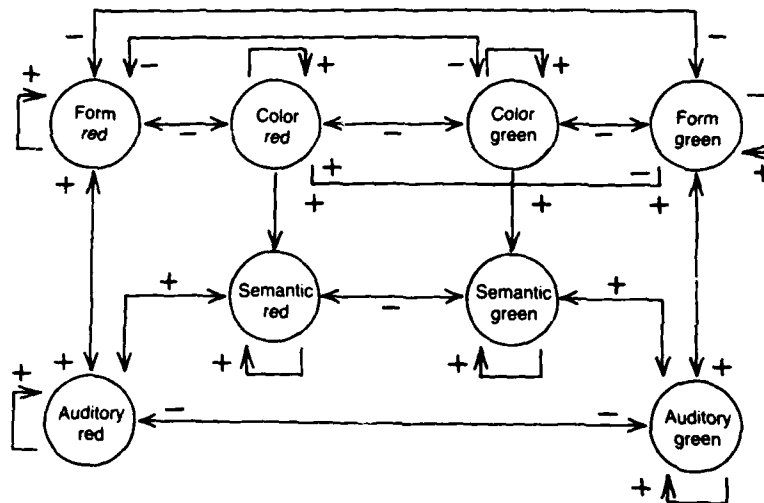
Attention and Problem Solving

Table 1

1. If S1 is on channel 1 -> do nothing

2. If S2 is on channel 1 -> Mark "target"

3. If S1 is on channel 2 -> do nothing

4. If S2 is on channel 2 -> Mark "target"

Production System for Visual Monitoring Study

----

84 04 11 050

Table 2

Visual Productions

If the instructions are to respond to the word and the word is
RED then -->
    Generate semantic code red .

If the instructions are to respond to the word and the word is
GREEN then -->
    Generate semantic code green .

If the instructions are to respond to the color and the color is
red then -->
    Generate semantic code red .

If the instructions are to respond to the color and the color is
green then -->
    Generate semantic code green .

Table 2 (continued)

Semantic Productions

If the semantic code is red then -->
  Generate auditory code /red/.

If the semantic code is green then -->
  Generate auditory code /green/.

Auditory Productions

If the auditory code is /red/ then -->
  Make external response "red."

If the auditory code is /green/ then -->
  Make external response "green."

Productions used to simulate Stroop Task.

END

DATE
FILMED

4 - 84

DTIC